

**What is claimed is:**

1           1.       A method comprising the steps of:  
2           receiving an email message having a word;  
3           generating a phonetic equivalent of the word;  
4           tokenizing the phonetic equivalent of the word to generate a token representative  
5 of the phonetic equivalent; and  
6           determining a spam probability from the generated token.

1           2.       The method of claim 1, wherein the step of generating the phonetic  
2 equivalent of the word comprises the steps of:  
3           identifying a string of characters, the string of characters including a non-  
4 alphabetic characters; and  
5           removing the non-alphabetic character from the string of characters.

1           3.       The method of claim 2, wherein the step of removing the non-alphabetic  
2 character comprises the step of:

3           locating a non-alphabetic character within the string of characters, the non-  
4 alphabetic character being at least one selected from the group consisting of:

5           " (quote);

6           ' (single quote);

7           ! (exclamation mark);

8           @ (at);

9           # (pound);

10          \$ (dollar);

11          % (percent);

12          ^ (caret);

13          & (ampersand);

14          \* (asterisk);

15          ( (open parenthesis);

16          ) (close parenthesis);

17          \_ (underscore);

18          - (hyphen);

19          + (plus);

20          = (equal);

21          \ (backslash);

22          / (slash);

23          ? (question mark);

24          (space);

25          (tab);

26 [ (open square bracket);  
27 ] (close square bracket);  
28 { (open bracket);  
29 } (close bracket);  
30 < (less than);  
31 > (greater than);  
32 , (comma);  
33 : (colon);  
34 ; (semi-colon);  
35 and . (period).

1 4. The method of claim 1, wherein the step of determining the spam  
2 probability comprises the steps of:  
3 assigning a spam probability value to the token; and  
4 generating a Bayesian probability value using the spam probability value assigned  
5 to the token.

1 5. The method of claim 4, wherein the step of determining the spam  
2 probability further comprises the step of:  
3 comparing the generated Bayesian probability value with a predefined threshold  
4 value.

1           6.       The method of claim 5, wherein the step of determining the spam  
2 probability further comprises the step of:  
3           categorizing the email message as spam in response to the Bayesian probability  
4 value being greater than the predefined threshold.

1           7.       The method of claim 5, wherein the step of determining the spam  
2 probability further comprises the step of:  
3           categorizing the email message as non-spam in response to the Bayesian  
4 probability value being not greater than the predefined threshold.

1           8.       A system comprising:  
2           means for receiving an email message having a word;  
3           means for generating a phonetic equivalent of the word;  
4           means for tokenizing the phonetic equivalent of the word to generate a token  
5 representative of the phonetic equivalent; and  
6           means for determining a spam probability from the generated token.

1           9.       A system comprising:  
2           receive logic configured to receive an email message having a word;  
3           phonetic logic configured to generate a phonetic equivalent of the word;  
4           tokenize logic configured to tokenize the phonetic equivalent of the word to  
5 generate a token representative of the phonetic equivalent; and  
6           spam-determination logic configured to determine a spam probability from the  
7 generated token.

1           10.     The system of claim 9, further comprising:  
2                 string-identification logic configured to identify a string of characters, the string of  
3 characters including a non-alphabetic characters; and  
4                 character-removal logic configured to remove the non-alphabetic character from  
5 the string of characters.

1           11.     The system of claim 10, further comprising:  
2                 spam-probability logic configured to assign a spam probability value to the token;  
3 and  
4                 Bayesian logic configured to generate a Bayesian probability value using the spam  
5 probability value assigned to the token.

1           12.     The system of claim 11, further comprising:  
2                 compare logic configured to compare the generated Bayesian probability value  
3 with a predefined threshold value.

1           13.     The system of claim 12, further comprising:  
2                 spam-categorization logic configured to categorize the email message as spam in  
3 response to the Bayesian probability value being greater than the predefined threshold.

1           14.     The system of claim 12, further comprising:  
2                 spam-categorization logic configured to categorize the email message as non-  
3 spam in response to the Bayesian probability value being not greater than the predefined  
4 threshold.

1           15.     A computer-readable medium comprising:  
2           computer-readable code adapted to instruct a programmable device to receive an  
3 email message having a word;  
4           computer-readable code adapted to instruct a programmable device to generate a  
5 phonetic equivalent of the word;  
6           computer-readable code adapted to instruct a programmable device to tokenize the  
7 phonetic equivalent of the word to generate a token representative of the phonetic  
8 equivalent; and  
9           computer-readable code adapted to instruct a programmable device to determine a  
10 spam probability from the generated token.

1           16.     The computer-readable medium of claim 15, further comprising:  
2           computer-readable code adapted to instruct a programmable device to identify a  
3 string of characters, the string of characters including a non-alphabetic characters; and  
4           computer-readable code adapted to instruct a programmable device to remove the  
5 non-alphabetic character from the string of characters.

1           17.     The computer-readable medium of claim 15, further comprising:  
2           computer-readable code adapted to instruct a programmable device to assign a  
3 spam probability value to the token; and  
4           computer-readable code adapted to instruct a programmable device to generate a  
5 Bayesian probability value using the spam probability value assigned to the token.

1           18.     The computer-readable medium of claim 17, further comprising:  
2           computer-readable code adapted to instruct a programmable device to compare the  
3     generated Bayesian probability value with a predefined threshold value.

1           19.     The computer-readable medium of claim 18, further comprising:  
2           computer-readable code adapted to instruct a programmable device to categorize  
3     the email message as spam in response to the Bayesian probability value being greater  
4     than the predefined threshold.

1           20.     The computer-readable medium of claim 18, further comprising:  
2           computer-readable code adapted to instruct a programmable device to categorize  
3     the email message as non-spam in response to the Bayesian probability value being not  
4     greater than the predefined threshold.